

New single day EM-seq and E5hmC-seq workflows: Faster library prep for whole genome and target enrichment methylome profiling



Daniel J. Evanich, Brittany S. Sexton, Robin Armstrong, Alexander Monovich, Vaishnavi Panchapakesa, Laura Blum, Ariel Erijman, Matthew A. Campbell, Bradley W. Langhorst, V K Chaithanya Ponnaluri, Louise Williams | New England Biolabs, Inc.

Introduction

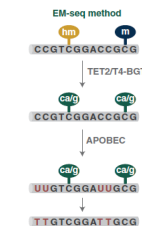
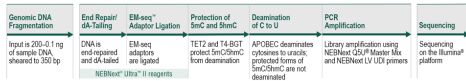
DNA methylation is an epigenetic regulator of gene expression with important functions in development and diseases such as cancer. Typically, the modified cytosines, 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), are detected by sequencing Illumina libraries generated using NEBNext® EM-seq™, an enzyme-based workflow, or bisulfite conversion. Alternatively, 5hmC can be specifically detected using another enzyme-based workflow called NEBNext E5hmC-seq™. In both EM-seq and E5hmC-seq, 5mC and/or 5hmC are protected during an initial enzymatic reaction from subsequent deamination by the enzyme APOBEC. Cytosines are deaminated to uracil in EM-seq while in E5hmC-seq, 5mC and cytosine bases are converted to thymine and uracil, respectively.

Library preparation time is important. Shorter workflows result in faster turn around times enabling increased sample processing and a more cost-effective library preparation. NEB has worked to improve library preparation time for EM-seq v2 and E5hmC-seq and has identified the deamination reaction as a potential target for time optimization. The current deamination time for EM-seq v2 and E5hmC-seq is 3 hours. Here we show compelling data that supports a 30 min deamination time for 100 ng to 0.1 ng DNA inputs. Furthermore, we demonstrate the effectiveness of combining NEBNext UltraShear®, a methylation friendly enzymatic fragmentation system, with EM-seq v2 and E5hmC-seq. This is an effective end-to-end library preparation solution that incorporates enzymatic fragmentation and does not require expensive instrument or consumables, as well as enabling automation. Additionally, as with EM-seq v1, EM-seq v2 is compatible with different enrichment strategies such as the Twist Human Methylome Panel.

The reduced deamination time makes the EM-seq v2 and E5hmC-seq workflow completion times comparable to current competitor methylation detection systems, for example, Watchmaker (TAPS+) and Illumina (5-base). The new deamination times for the EM-seq v2 and E5hmC-seq protocols enables a comfortable single day workflow resulting in faster turn around time leading to higher throughput without compromising the high quality of data that EM-seq is known for.

Methods

EM-seq v2: Enzymatic Detection of 5mC & 5hmC



EM-seq v2 Library Construction:

- 200, 100, 10, 1 & 0.1 ng of NA12878 genomic DNA spiked with control DNAs: unmethylated lambda DNA and CpG methylated pUC19.
- DNA was fragmented using Covaris or NEBNext UltraShear.
- Fragmented DNA was end-repaired then ligated to EM-seq adaptors.
- 5mC and 5hmC were protected using TET2 and T4-BGT from deamination by APOBEC (3 hours or 30 min).
- Libraries were sequenced on an Illumina NovaSeq 6000 using 2x 150 base reads. 800 million total reads were analyzed per library.

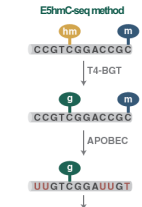
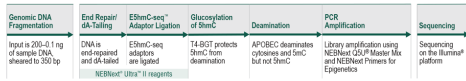
Hybrid Selection of EM-seq v2 Libraries using Twist Methylome Panel



Hybrid Capture of EM-seq v2 Libraries:

- 200, 100, 10 & 1 ng of NA12878 genomic DNA were spiked with control DNAs: unmethylated lambda DNA and CpG methylated pUC19.
- DNA was fragmented using Covaris or NEBNext UltraShear.
- EM-seq v2 libraries were made using either a 3 hour or a 30 min deamination reaction. Target capture libraries were amplified at 10 and 1 ng inputs for either 3 additional cycles for Covaris libraries or 2 additional cycles for UltraShear libraries compared to whole genome libraries. 100 ng DNA inputs used 2 extra PCR cycles.
- Twist's Human Methylome Panel was used to perform target capture following the manufacturer's recommendations.
- Libraries were sequenced on an Illumina NovaSeq 6000 using 2x 150 base reads and 200 million total reads were analyzed per target capture library.
- Matched cycle whole genome and target capture libraries were used in analysis.

E5hmC-seq: Enzymatic Detection of 5hmC



E5hmC-seq Library Construction:

- 200, 100, 10, 1 & 0.1 ng of human brain genomic DNA spiked with control DNAs: unmethylated lambda DNA and T4 phage DNA (all cytosines are 5hmC modified).
- DNA was fragmented using Covaris or NEBNext UltraShear.
- Fragmented DNA was end-repaired then ligated to E5hmC-seq adaptors.
- 5hmCs were protected using T4-BGT from deamination by APOBEC (3 hours or 30 min).
- Libraries were sequenced on an Illumina NovaSeq 6000 using 2x 150 base reads.

Sequencing and Data Analysis

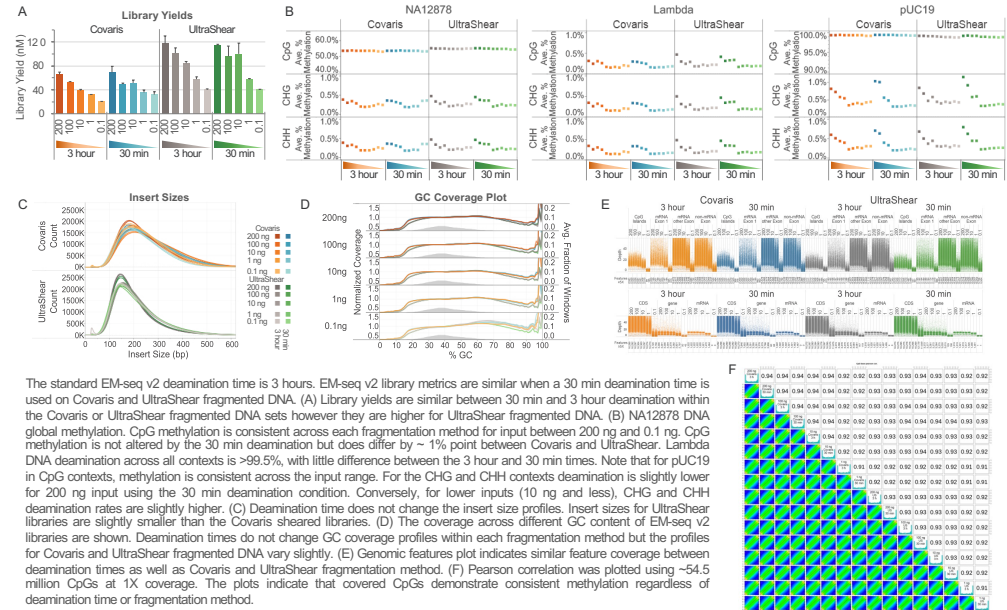
EM-seq v2, Target capture and E5hmC-seq data analysis



- Reads were adaptor trimmed (fastp) then aligned to a human T2T composite genome (including controls) using bwa-meth.
- 5mC and 5hmC information was extracted from the alignments using MethlyDackel and levels were evaluated independently for each chromosome.
- methylKit data was used for Pearson correlation at 1x minimum coverage.
- Picard was used to mark duplicates as well as calculate library insert size and GC bias.
- Target enrichment Data was analyzed as for EM-seq and in addition, Picard was used to calculate HS metrics.

Results

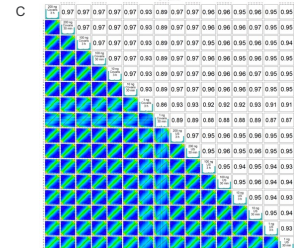
EM-seq v2 library construction: A single day sample to library workflow



EM-seq v2 Target Capture

PCR Cycles	Covaris		UltraShear	
	3 hours	30 min	3 hours	30 min
DNA input (ng)	200	4	4	4
100	7	7	7	7
10	11	11	10	10
1	14	14	13	13
Average PCR Yield	200	137.8	161.2	185.6
100	459.9	667.8	556.2	724.5
10	328.1	362.1	374.0	457.3
1	181.4	133.9	257.6	288.2

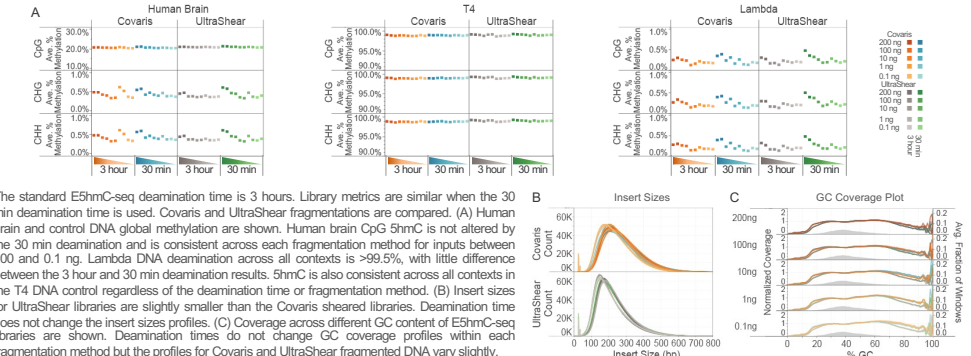
Note: 115 ng of each EM-seq v2 library went into target capture



	Covaris		UltraShear	
	3 hours	30 min	3 hours	30 min
Mean Target Coverage	200: 87.0X, 100: 88.0X, 10: 59.0X, 1: 14.0X	200: 91.0X, 100: 69.0X, 10: 11.0X, 1: 14.0X	200: 89.0X, 100: 68.0X, 10: 8.0X, 1: 11.0X	200: 92.0X, 100: 88.0X, 10: 12.0X, 1: 22.0X
Target Bases > 30X	200: 98%, 100: 95%, 10: 95%, 1: 0%	200: 98%, 100: 97%, 10: 97%, 1: 0%	200: 98%, 100: 97%, 10: 97%, 1: 0%	200: 98%, 100: 97%, 10: 97%, 1: 0%
Fold 80 base penalty	200: 1.4, 100: 1.4, 10: 1.3, 1: 1.3	200: 1.4, 100: 1.4, 10: 1.3, 1: 1.3	200: 1.4, 100: 1.4, 10: 1.4, 1: 1.5	200: 1.4, 100: 1.4, 10: 1.4, 1: 1.5
% zero coverage targets	200: 0.5, 100: 0.5, 10: 0.5, 1: 0.6	200: 0.5, 100: 0.5, 10: 0.5, 1: 0.6	200: 0.5, 100: 0.5, 10: 0.5, 1: 0.6	200: 0.5, 100: 0.5, 10: 0.5, 1: 0.5

EM-seq v2 libraries made using either a 3 hour or 30 min deamination time were used to make hybrid capture libraries. (A) Table showing PCR cycles and average EM-seq v2 library yields. Library yields are higher for UltraShear fragmented DNA with either same (200 ng and 100 ng) or one less (10 ng and 1 ng) PCR cycles compared to Covaris fragmented DNA. Deamination time did not affect yields within a fragmentation method. (B) Hybrid selection metrics. The 200 ng and 100 ng inputs performed most robustly with high Mean Target Coverage (MTC), (~80X - 90X), the 10 ng inputs had a MTC of ~60X - 70X and the 1 ng input had a MTC of ~10X - 20X. Target Bases > 30X were >90% except for 1 ng inputs. Covaris performed poorly at 0% and UltraShear libraries were around 20%. All libraries performed similarly for Fold 80 base penalty and % zero coverage targets. (C) Pearson correlation was plotted using ~14.1 million CpGs. The plots indicate covered CpGs demonstrate consistent methylation regardless of deamination time or fragmentation method. Please see NEB Suite poster "EM-seq enables simultaneous analysis of methylation and germline variants from whole genome and target capture data" for information on variant calling using target enriched libraries.

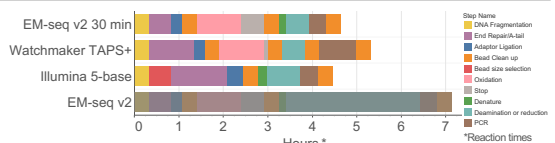
E5hmC-seq libraries made using the single day sample to library workflow



Conclusions

Single day EM-seq v2 & E5hmC-seq workflows

- 30 min deamination gives robust metrics across wide input range, independent of fragmentation method.
- End-to-end enzymatic solution for methylome analysis: UltraShear upstream of EM-seq v2.
- Targeted cost effective methylomes: combining EM-seq v2 with hybrid capture.



Acknowledgement

Authors would like to acknowledge the technical assistance provided by Dora Postal, Kristen Augulewicz, Harry Bell, and Rebecca Gawron at the New England Biolabs Sequencing Core Facility.